

• Supplementary File •

# A Nonprofiled Side-Channel Analysis Based on Variational Lower Bound Related to Mutual Information

Chi ZHANG<sup>1</sup>, Xiangjun LU<sup>1</sup>, Pei CAO<sup>1</sup>, Dawu GU<sup>1\*</sup>, Zheng GUO<sup>2</sup> & Sen XU<sup>3</sup>

<sup>1</sup>*School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China;*

<sup>2</sup>*ZhiXun Crypto Testing and Evaluation Technology Co., Ltd., Shanghai 201601, China;*

<sup>3</sup>*Viewsource Information Science and Technology Co., Ltd., Shanghai 200241, China*

## Appendix A Comparison of Hyperparameters on the High-SNR Dataset.

Figure A1 and Figure A2 illustrate the GEs of NMIA and NNSCA for all possible values of the hyperparameters on the high-SNR dataset, respectively. The x-axis of each subfigure is the number of traces  $n_t$ , and the y-axis of each subfigure is the GE. Each column corresponds to one specific  $n_{\text{mlp}}$ , and each row corresponds to one specific  $l$  and  $\sigma_{\text{mlp}}$ . For LReLU and ELU, five possible values of  $\alpha$ , i.e.,  $\alpha \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$  are presented by different colors and different line styles shown in the legend.

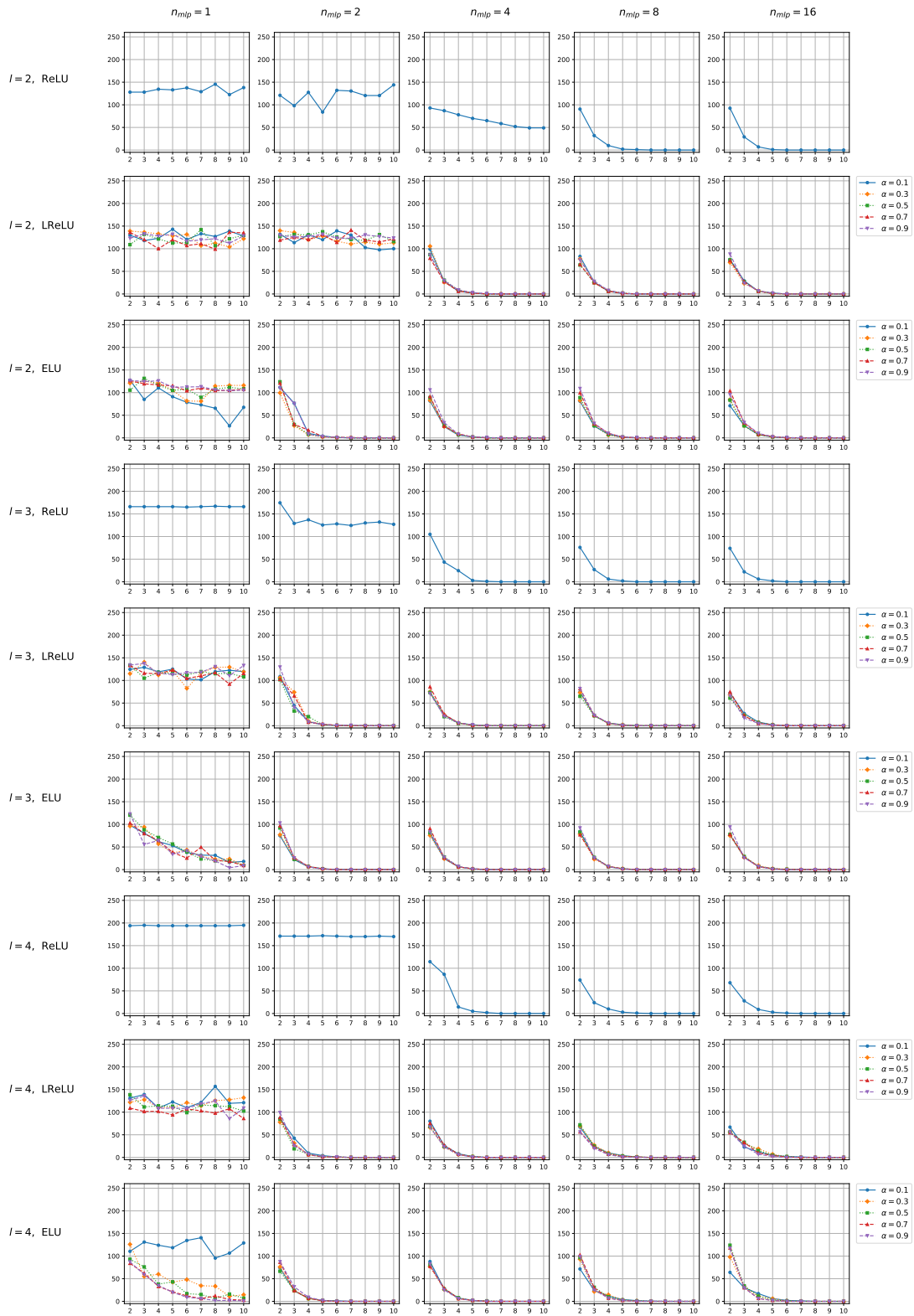
For NMIA, when  $n_{\text{mlp}} = 1$  and  $l \in \{2, 3, 4\}$ , or  $n_{\text{mlp}} = 2$  and  $l = 2$ , the MLP with ELU has better performance for all  $l \in \{2, 3, 4\}$ ; for the other combinations of  $n_{\text{mlp}}$  and  $l$ , the MLP with LReLU and the MLP with ELU have the similar performance. When  $n_{\text{mlp}} \geq 4$ , the values of  $l$  and  $\alpha$  have little influence on the position that the GE becomes 0. When  $n_{\text{mlp}} < 4$ , larger  $l$  leads to better GE.

For NNSCA, LReLU and ELU are also better than ReLU. When  $n_{\text{mlp}} \geq 2$ , the values of  $l$  and  $\alpha$  have little influence on the GE. When  $n_{\text{mlp}} = 1$ , the MLP with LReLU behave more stably for all  $l \in \{2, 3, 4\}$ .

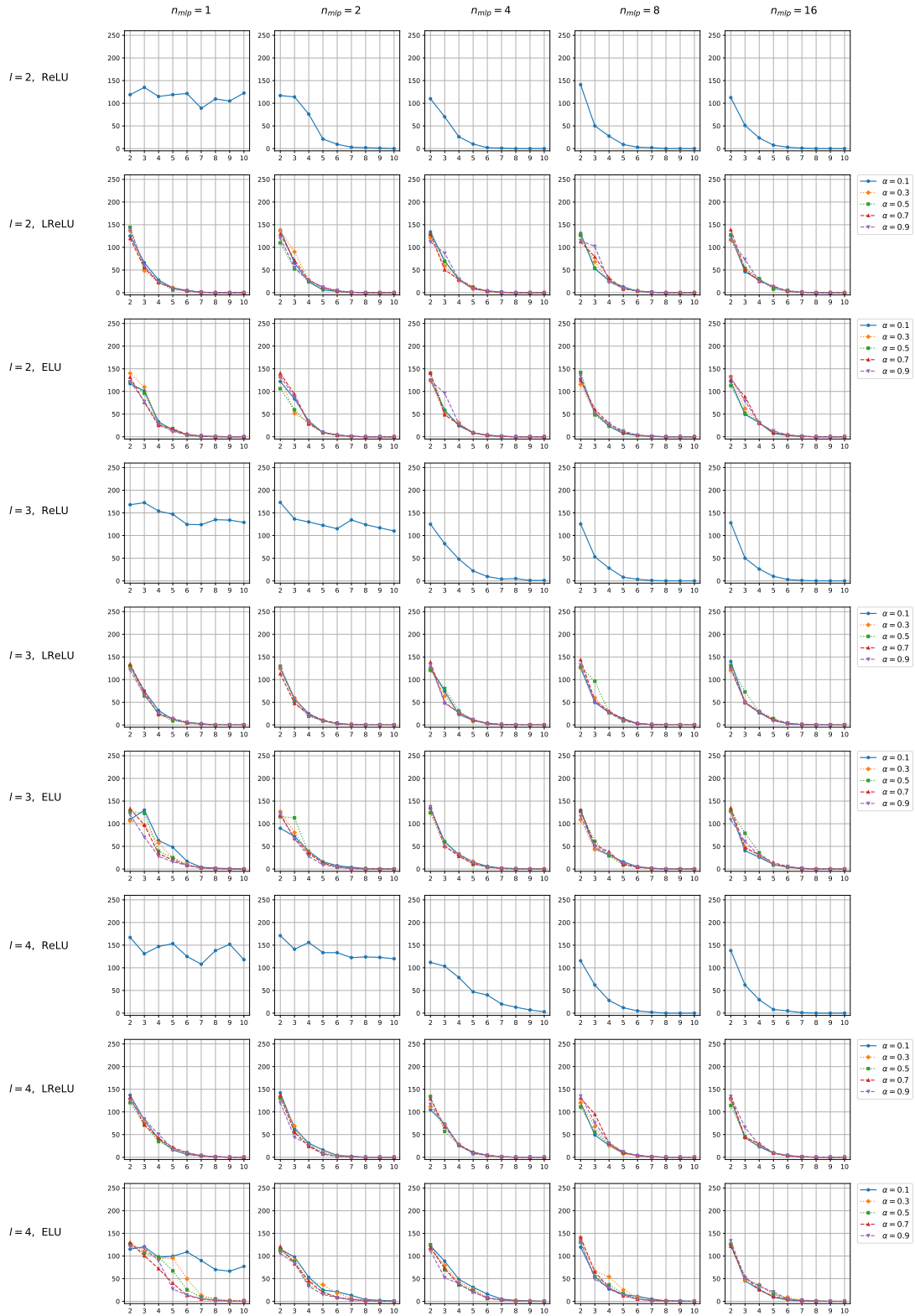
In summary, we regard LReLU as the appropriate activation function in our case because the computational burden of LReLU is less than that of ELU. The LReLU with  $\alpha = 0.9$  behaves well on average. In addition, the ranges of  $n_{\text{mlp}}$  and  $l$  we defined are reasonable, and we can use them in the experiments.

---

\* Corresponding author (email: dwgu@sjtu.edu.cn)



**Figure A1** The GE of NMIA on high-SNR dataset for all possible hyperparameters.



**Figure A2** The GEs of NNSCA on high-SNR dataset for all possible hyperparameters.